



# 75th MORSS CD Cover Slide

712CD

For office use only 41205

UNCLASSIFIED DISCLOSURE FORM CD  
Presentation

12-14 June 2007, at US Naval Academy, Annapolis, MD

**Please complete this form 712CD as your cover page  
to your electronic briefing submission to the MORSS  
CD. Do not fax to the MORS office.**

**Author's Request: I request authority to disclose the following  
presentation in the MORSS Final Report, for inclusion on the MORSS CD  
and/or posting on the MORS Website.**

Name of Principal Author and all other author(s): Gregory T. Hutto, YD-3 DAF

Principal Author's Organization and address: Phone: \_850,882,0607\_

Fax: \_850.882.5644\_

Email: \_gregory.hutto@eglin.af.mil\_

Original title on 712 A/B: \_Resampling Statistics\_

Revised title: \_\_\_\_\_

Presented in: ***Tutorial Session***

This presentation is believed to be:  
Distribution A: UNCLASSIFIED AND  
APPROVED FOR PUBLIC RELEASE

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>01 JUN 2007</b>		2. REPORT TYPE <b>N/A</b>		3. DATES COVERED <b>-</b>	
4. TITLE AND SUBTITLE <b>Resampling Statistics Tutorial</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>53rd Wing Air Combat Command USAF Eglin AFB</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release, distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>See also ADM202526. Military Operations Research Society Symposium (75th) Held in Annapolis, Maryland on June 12-14, 2007, The original document contains color images.</b>					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>UU</b>	18. NUMBER OF PAGES <b>45</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			



Baron Von Munchhausen

# Resampling Statistics Tutorial

## 75<sup>th</sup> MORS Symposium USNA

---

Presented by:

Greg Hutto  
Chief Operations Analyst  
53<sup>rd</sup> Wing  
Air Combat Command USAF  
[gregory.hutto@eglin.af.mil](mailto:gregory.hutto@eglin.af.mil)



Baron Von Munchhausen

# Tutorial Overview

- ☐ A couple of motivational problems
- ☐ Why we care & Resampling History
- ☐ Resampling vs. Classical Approach to Statistical Testing
  - The one sample bootstrap (binomial p to 1 sample t))
  - The two sample bootstrap & shuffle (two sample t)
  - One sample – sample size (noncentral t power)
  - 1 Way ANOVA (F test)
  - Simple linear regression
  - 2 Way ANOVA
  - Medians, correlations, percentiles
- ☐ Potential Advantages of Adopting Resampling
  - Teaching
  - Dealing with missing data
  - Estimating sample size
  - Nonnormal data
  - Getting the “right” answer in Analysis
- ☐ Caveats & Directions for the Future



Baron Von Munchhausen

## Why we care about resampling

- ☐ It's all about test productivity .., how many *good* tests can you run with \$48 million and 2200 people?
- ☐ As of Spring 2002, the 53<sup>rd</sup> Wing has changed its method of test to Design of Experiments (DOE)
- ☐ We have 350+ project officers and 70 analysts to teach on a rotating basis
- ☐ More than 20 geographically separated units
- ☐ As of last count, 256 test projects on the ACC Test Priority List (March 2007)
- ☐ We currently service less than 75% of these each year
- ☐ And .. Post-Sept 11, the pace continues to quicken



Baron Von Munchhausen

# Problems to motivate usefulness of resampling

---

- ☐ Probability of 1 boy & 1 girl in a two child family?
- ☐ Given that one child is a boy in a two child family, probability the other child is also a boy?
- ☐ JASSM Reliability is low. If “fixed” and now 80% can we determine that with 20 shots?
- ☐ Joint Programmable Fuze is 90%+ reliable. Or is it? We had 2 or 3 of 6 failures.
- ☐ The SWEAT diet is being proposed to get our Airmen “Fit to Fight” will it work?



Baron Von Munchhausen

# Why Resample? Information gathering & data analysis

---

- ☐ Answering a research question
- ☐ Estimating an uncertain quantity
- ☐ Determine sampling distributions of statistics whose distributions cannot or have not been mathematically approximated
- ☐ To avoid doubtful assumptions about the data
- ☐ In the case of permutation tests, to arrive at exact p-values
- ☐ To use a conceptually simple, widely applicable general method



Baron Von Munchhausen

## Caveat – Resampling is a work in progress in our Wing

---

- ☐ Col (Dr) Pete Vandenbosch suggested we consider resampling in January 02
- ☐ Used 2002 MORS Symposium as a forcing function to get to work on solving the problems of applying
- ☐ Made some progress – Overall, a B+
  - Teaching – A
  - Nonnormal data – B+
  - Missing data – B-
  - Sample Size – A
  - Applying to DOE – B
- ☐ More directions at the summary



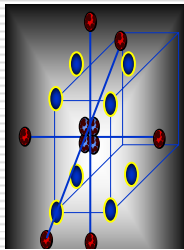


# Teaching Difficult Statistical Concepts via Resampling



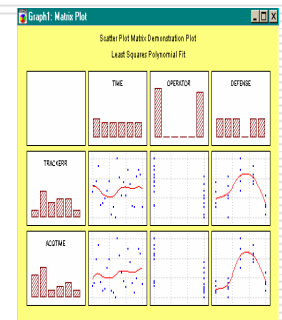
Baron Von Munchhausen

ECM	Offset	Clutter	High Altitude		Low Altitude	
			Auto Trk	Man Trk	Auto Trk	Man Trk
Wet	Short	High	12.1			5.6
		Low		14.7	4.6	
	Long	High		12.2	7.8	
		Low	13.6			3.4
Dry	Short	High		12.1	5.6	
		Low	14.7			4.6
	Long	High	12.2			7.8
		Low		13.6	3.4	



Central Composite

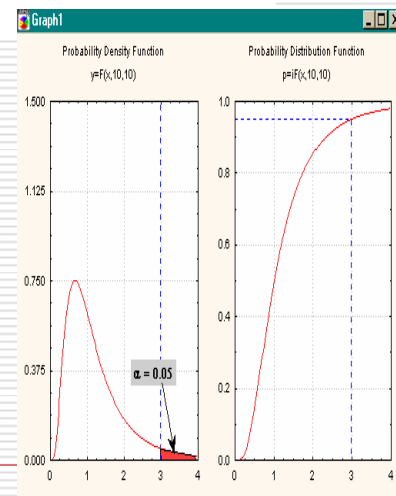
- "Statistically Significant Difference"
- Hypothesis testing
  - $\alpha$  and  $\beta$  errors
  - Confidence intervals
  - Sample size via OC curves
  - Random va. Causal variation
  - Regression coefficients



ANOVA; Var.:LOCERRPC; R-sqr=.37782; Adj:2285					
EXPERIM. DESIGN	2**(3-0) design; MS Residual=10450.88 DV: LOCERRPC				
Factor	SS	df	MS	F	p
(1)THREAT	39466.7	1	39466.67	3.776398	.063321
(2)BOUND	7602.1	1	7602.08	.727411	.401821
(3)JINK	32864.8	1	32864.76	3.144689	.088359
1 by 2	61.6	1	61.56	.005890	.939436
1 by 3	78618.9	1	78618.86	7.522705	.011100
2 by 3	45.0	1	45.02	.004307	.948194
Error	261271.9	25	10450.88		
Total SS	419930.8	31			

June 2007

Resampling Statistics Tutorial



R-7



# Resampling crystallizes thinking



Baron Von Munchhausen

- ☐ Why must  $\alpha$  be set to compute  $\beta$  error?
- ☐ What is the root problem behind missing data and unbalanced designs?
- ☐ What assumptions are we making about the sample when we apply the t test? Are they appropriate?
- ☐ What are the real effects of outliers on linear models?
- ☐ What exactly are my null and alternate worlds?



Baron Von Munchhausen

# Definitions

---

**Bootstrap:** Statistics done under the assumption that the population distribution is identical to the sample distribution. The data are sampled with replacement, simulating an infinite population from the sample.

**Permutation (Shuffling) Methods:** The basis for Fisher's Exact test and the Tea Experiment. The gold standard of resampling schemes – sampling is without replacement.

**Resampling:** Statistics done under computation-intensive methods involving sampling from some distribution. These include bootstrap, jackknife, data shuffling schemes, and possibly others.

**Monte Carlo:** Simulations done using random (or pseudo-random) numbers. These might include resampling schemes, but for our purposes today, we'll distinguish between them.



# Historical Resampling Timeline<sup>1</sup>



Baron Von Munchhausen

1650

Gambling Experiments – Blasé Pascal

1700

Probability Theory – Bernoulli

1850

1900

Student t Distribution W.T. Gosset

The Tea Experiment – R.A. Fisher

1925

Fisher's Exact Test (permutations) – R.A. Fisher

1950

Monte Carlo approaches (RAND et al.)

Quenouille's (and later Tukey's) Jackknife

1975

Applications to Business and Economics – Julian Simon

1980

Publication of Efron's (Stanford) Article on Bootstrap

1990

Peter Hall's Publication of Asymptotic Theory of Resampling

2000

1. Source: Chapter 1 *Bootstrap Methods and Their Application*, AC Davidson, 1997



# Is Bootstrap a Fraud?



Baron Von Munchhausen



THE TRAVELS  
AND  
SURPRISING ADVENTURES  
OF  
BARON MUNCHAUSEN  
ILLUSTRATED WITH  
*THIRTY-SEVEN*  
*CURIOUS ENGRAVINGS*

FROM  
THE BARON'S OWN DESIGNS  
AND FIVE ILLUSTRATIONS  
By G. CRUIKSHANK

**“Much to my chagrin, I found myself at the bottom of the lake!” declared the Baron.**

**“But Baron, how did you save yourself from a watery fate?” exclaimed his lady listener.**

**“Why, I simply reached down,” the Baron explained “ and pulled myself up by my bootstraps.”**

Source: [www.munchasuen.com](http://www.munchasuen.com)



# Gosset Resampled to Generate the Student t Distribution



VOLUME VI

MARCH, 1908

No. 1

## BIOMETRIKA.

---

### THE PROBABLE ERROR OF A MEAN.

By STUDENT.

#### *Introduction.*

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

- ☐ Article originating the Student t distribution by W.T. Gosset, a chemist for Guinness Brewery
- ☐ Wrote under a pseudonym for the same reason Romance Writers do ... Statistics not an honorable profession





Baron Von Munchhausen

# Gosset's 1908 Article

## THE PROBABLE ERROR OF A MEAN.

By STUDENT.

### *Introduction.*

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty:—(1) owing to the "error of random sampling" the mean of our series of experiments deviates more or less widely from the mean of the population, and (2) the sample is not sufficiently large to determine what is the law of distribution of individuals. It is usual, however, to assume a normal distribution, because, in a very large number of cases, this gives an approximation so close that a small sample will give no real information as to the manner in which the population deviates from normality: since some law of distribution must be assumed it is better to work with a curve whose area and ordinates are tabled, and whose

Translation – we use the normal because we know it...  
"When all you have is a hammer, everything begins to look like a nail."



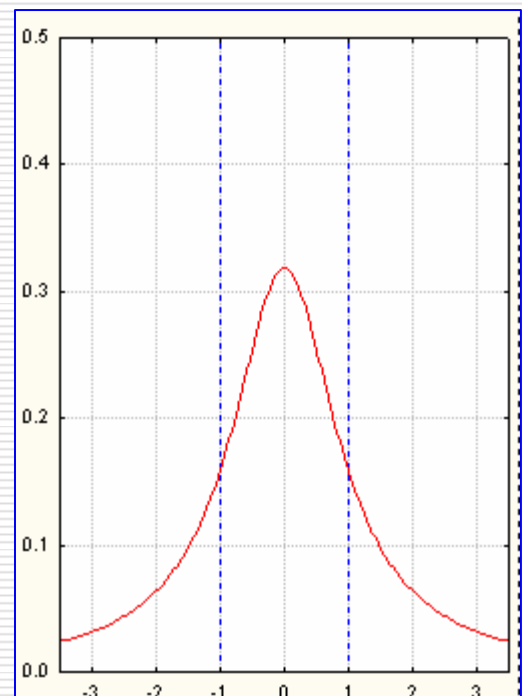
Baron Von Munchhausen

## Conclusion ...

### SECTION VI. *Practical Test of the foregoing Equations.*

Before I had succeeded in solving my problem analytically, I had endeavoured to do so empirically. The material used was a correlation table containing the height and left middle finger measurements of 3000 criminals, from a paper by W. R. Macdonell (*Biometrika*, Vol. I. p. 219). The measurements were written out on 3000 pieces of cardboard, which were then very thoroughly shuffled and drawn at random. As each card was drawn its numbers were written down in a book which thus contains the measurements of 3000 criminals in a random order. Finally each consecutive set of 4 was taken as a sample—750 in all—and the mean, standard deviation, and correlation† of each sample determined.

- ☐ Simulating a Shuffle test – *without* replacement
- ☐ Could have improved his procedure with bootstrap, I think, and by repeating the sets of 750 -- 3 or 4 times.







# Fisher Shuffled for his Randomization (Exact) Test

*In this clever experiment, Fisher justified the  $t$  and  $F$  statistics to examine the null hypothesis of sameness, without appealing to the Central Limit Theorem or the Chi Square Theorem. It makes us more comfortable with violations of the normality and equality of variance assumptions.*

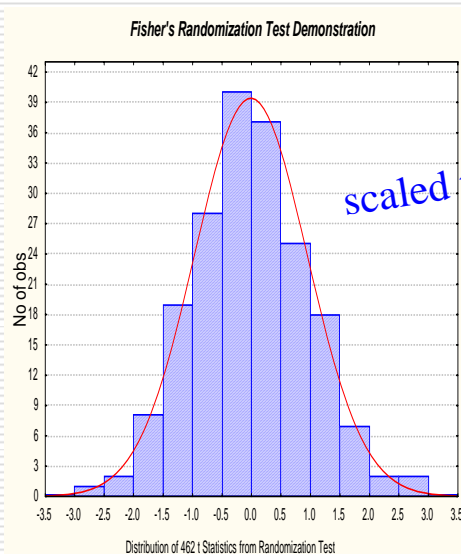
Ordering 1		Ordering 2	
Wet	Dry	Wet	Dry
30	26	30	26
11	24	11	24
25	28	25	28
16	14	18	14
21	18	21	16
	25		25

$n_W = 5$     $n_D = 6$   
 $\bar{y}_W = 21$     $\bar{y}_D = 23$   
 $\Delta y = 2$

$H_0 \quad \mu_W = \mu_D$   
 $H_a \quad \mu_W \neq \mu_D$

$$\begin{bmatrix} 11 \\ 5 \end{bmatrix} = \frac{11!}{5!6!} = 462$$

These are two of 462 ways to assign 5 samples to the “Wet” label. Under  $H_0$ , (W=D), the W/D labels make no difference in  $y$  values and thus the samples randomly appear in each column. We plot the distribution of mean differences, and observe how often a difference in 2 of means appears. This is the *randomization test*



In this problem, the alphas for both the  $t$  and the randomization test agree:

$$\alpha_t = .34$$

$$\alpha_{\text{RAND}} = .33$$

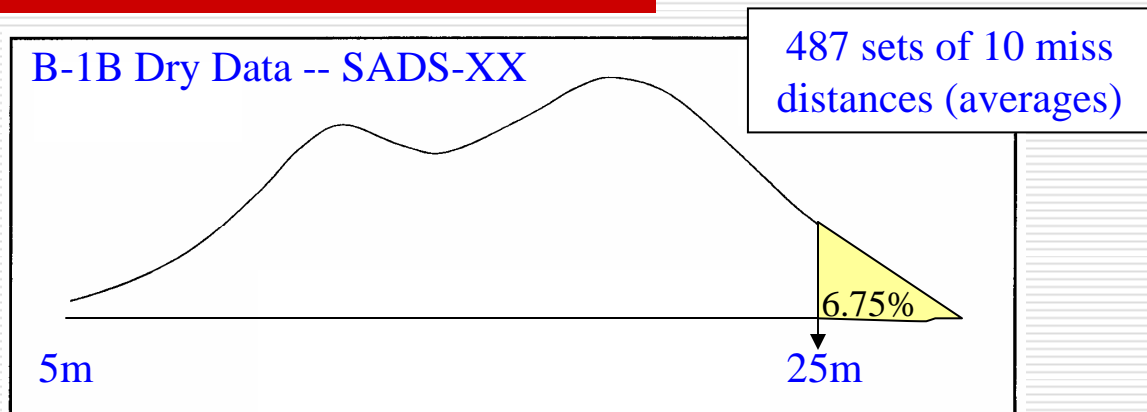
**Both  $t$  and  $F$  can be shown to fit this distribution . . .**



# Concept of an External Reference Distribution and Alpha Error



Baron Von Munchhausen



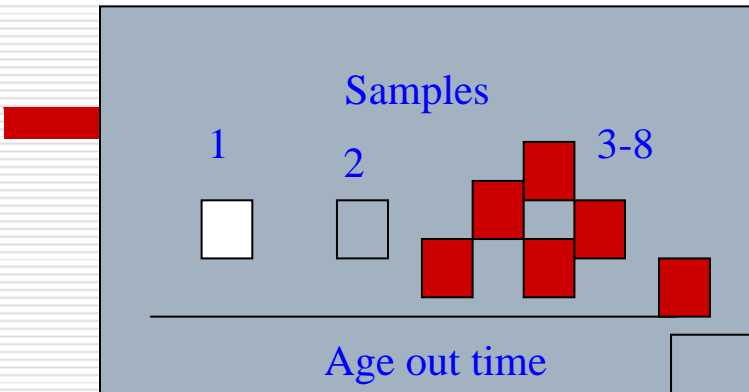
- ☐ Suppose I am a data collector - only the best! I have historical records of B-1B dry data sets
- ☐ Someone offers me a data set claiming it's B1-B dry data.  $x = 25m$ ,  $n = 10$
- ☐ Examining my reference collection, I observe that only 6.75% of my 487 samples equal or exceed this value
- ☐ Accordingly, I reject this set with a 6.75% chance of being mistaken
- ☐ I find likelihood that I see a mean of 25m given this is B-1B dry is low



# Developing a Random Variable Distribution

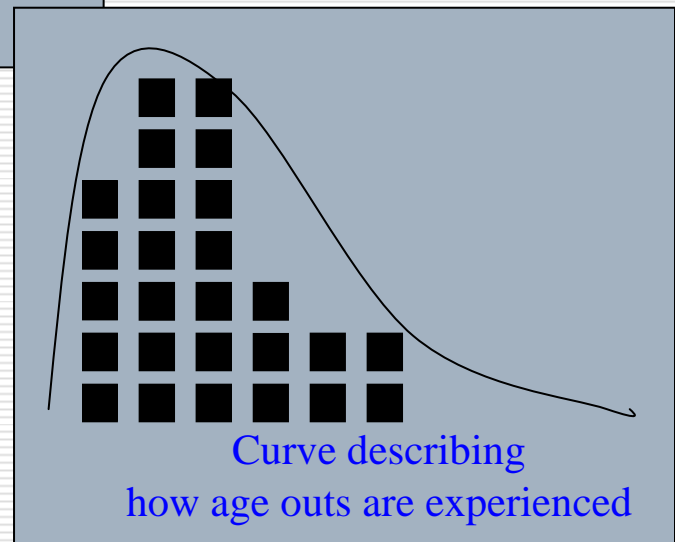


Baron Von Munchhausen



*Step 1. Collect some data*

*Step 2. Assess a discrete or continuous curve that describes how samples are observed*



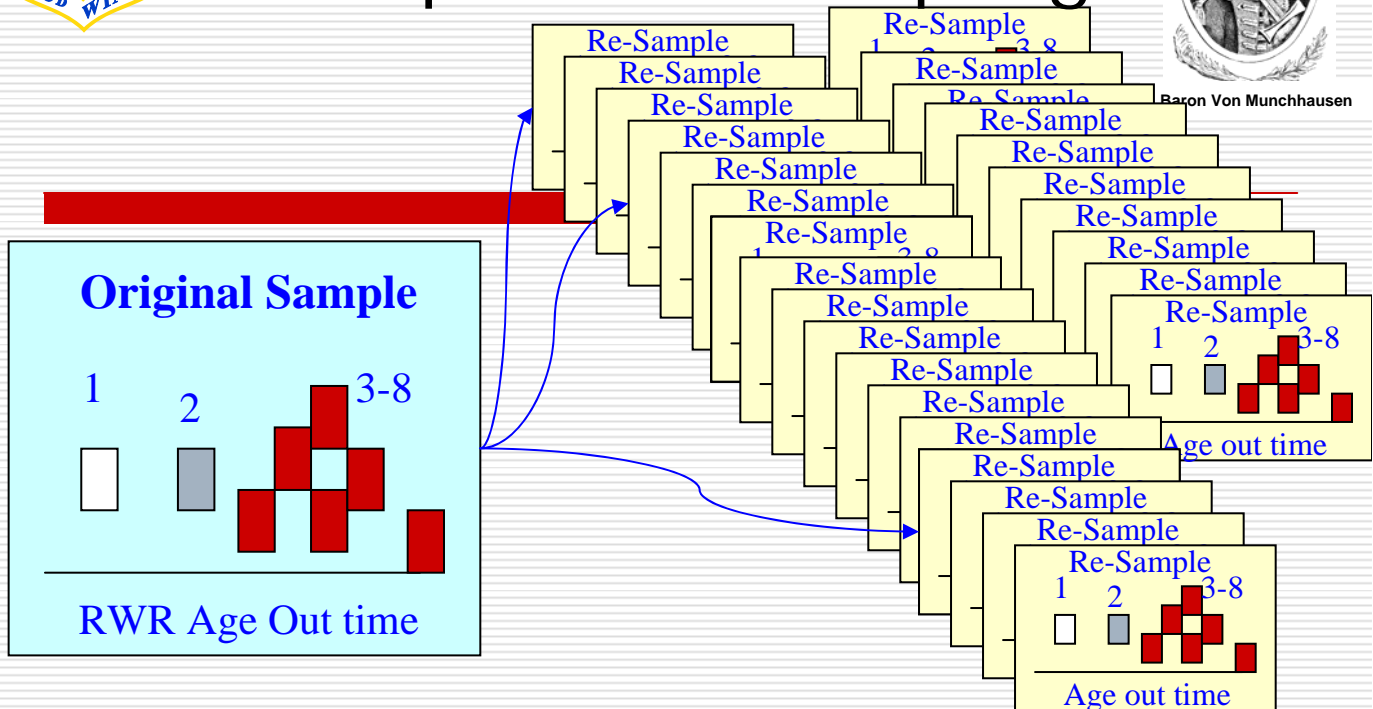
- ❑ A random variable converts experimental outcomes into a numeric variable that can be discussed with the tools we are learning.
- ❑ What proportion of RWR age outs are greater than 12 seconds?
- ❑ With a reference distribution you can compute "unlikelihood"
- ❑ **Probability distribution:** Features  $P(x)$  values for each outcome.  $P(x)$  are between zero and one. The  $p$ 's sum to 1.0
- ❑ We have several classical theoretical ones: binomial, t, z, chi-square, F ...



# Concepts of Resampling



Baron Von Munchhausen



- ❑ As with classical statistical theory – what we know of the population is contained in the sample
- ❑ With classical stats, we use mathematical theorems to estimate the behavior of the population
- ❑ With resampling – we draw repeated samples from our original sample and use that evidence to describe the population
  - Conceptually easier to swallow (and teach)
  - Fewer distributional assumptions
  - More flexible in practice
- ❑ Let's see where this idea takes us...



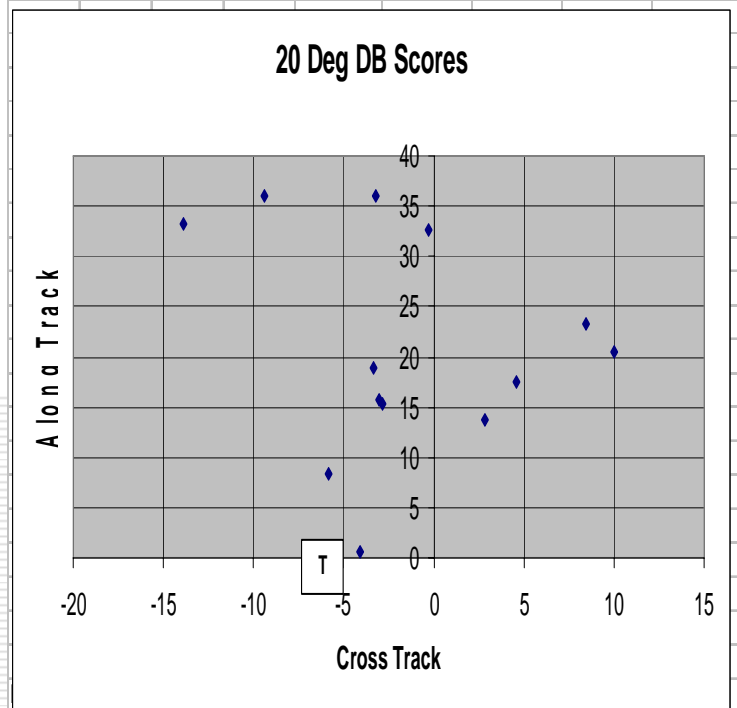
# As we re-sample, we obtain different values for a Random Variable



Baron Von Munchhausen

- Original sample of 13 weapon scores
- Resample in groups of 13 (with replacement)

20 Degree Dive Bomb Bomb Scores				Sampling Distributions of Random Variable			
Sample	X	Y	ReSample	Mean Y	Med Y	SD Y	X(90)
1	-4	1	1	33.2	36.0	17.5	15.8
2	-3	36	2	15.2	18.9	20.5	36.0
3	-9	36	3	33.2	33.2	0.7	33.2
4	3	14	4	15.2	33.2	32.7	23.3
5	-3	19	5	15.8	32.7	15.2	17.5
6	10	20	6	17.5	15.8	33.2	0.7
7	-3	15	7	17.5	18.9	32.7	35.9
8	8	23	8	32.7	23.3	15.2	15.2
9	0	33	9	0.7	36.0	15.8	15.2
10	-14	33	10	15.2	33.2	13.7	13.7
11	-3	16					
12	5	18					
13	-6	8					
Resample				19.6	32.9	10.4	36.0
Original				22.6	18.9	11.0	35.4

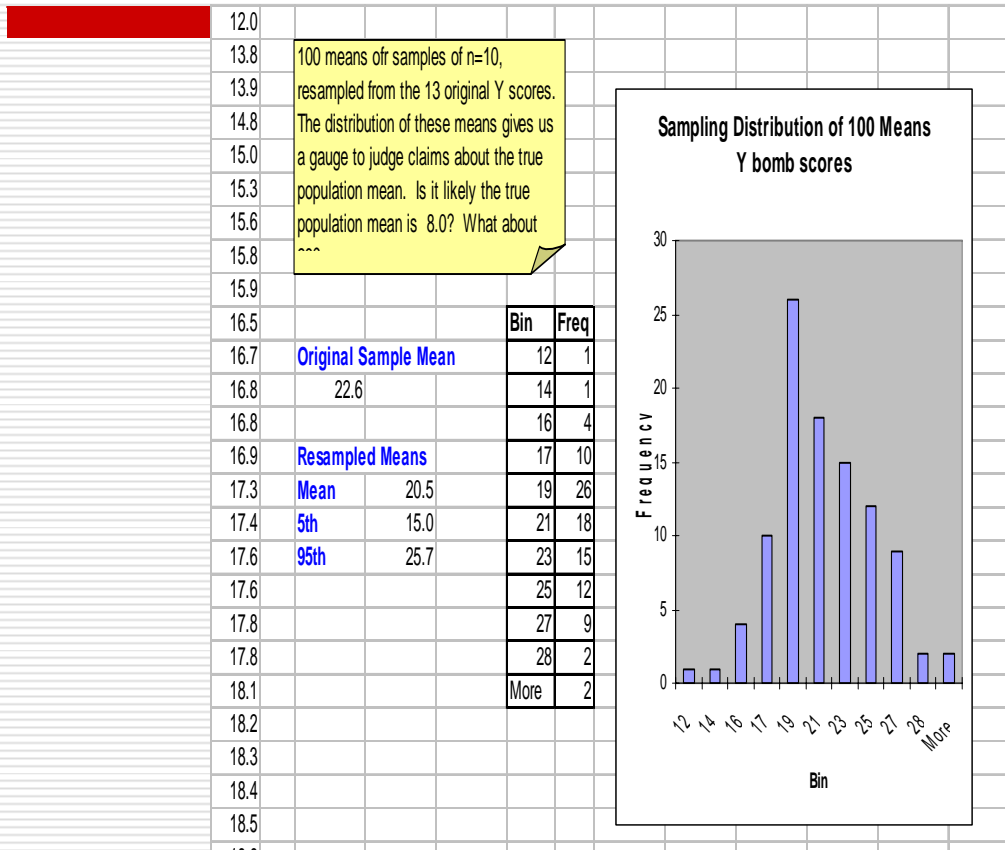




# Distribution of the Mean as we re-sample (sampling distribution)



Baron Von Munchhausen



- ☐ Resampling statistics (from Simon '68 and Efron '79) supply us with that empirical reference distribution we had been hoping for
- ☐ We can make probability statements regarding the behavior of any function computed from the sample



# Central Limit Theorem -- Recap



Baron Von Munchhausen

## Given:

1. Random variable  $x$  distributed arbitrarily with mean  $\mu$  and standard deviation  $\sigma$ .
2. Draw samples of size  $n$  are randomly (independently!) from this population.

## Conclusions:

1. The distribution of sample means  $\bar{X}$  will, as the sample size increases, approach normal.
2. The mean of the sample means will be the population mean  $\mu$ .
3. The standard deviation of the sample means is  $\frac{\sigma}{\sqrt{n}}$  (the standard error of the mean)

## Practical Rules of Thumb

1. For samples of size  $n$  larger than 30, the distribution of the sample means can be approximated reasonably well by a normal distribution. The approximation gets better as the sample size  $n$  becomes larger.
2. If the original population is itself normally distributed, then the sample means will be normally distributed for any sample size  $n$ .

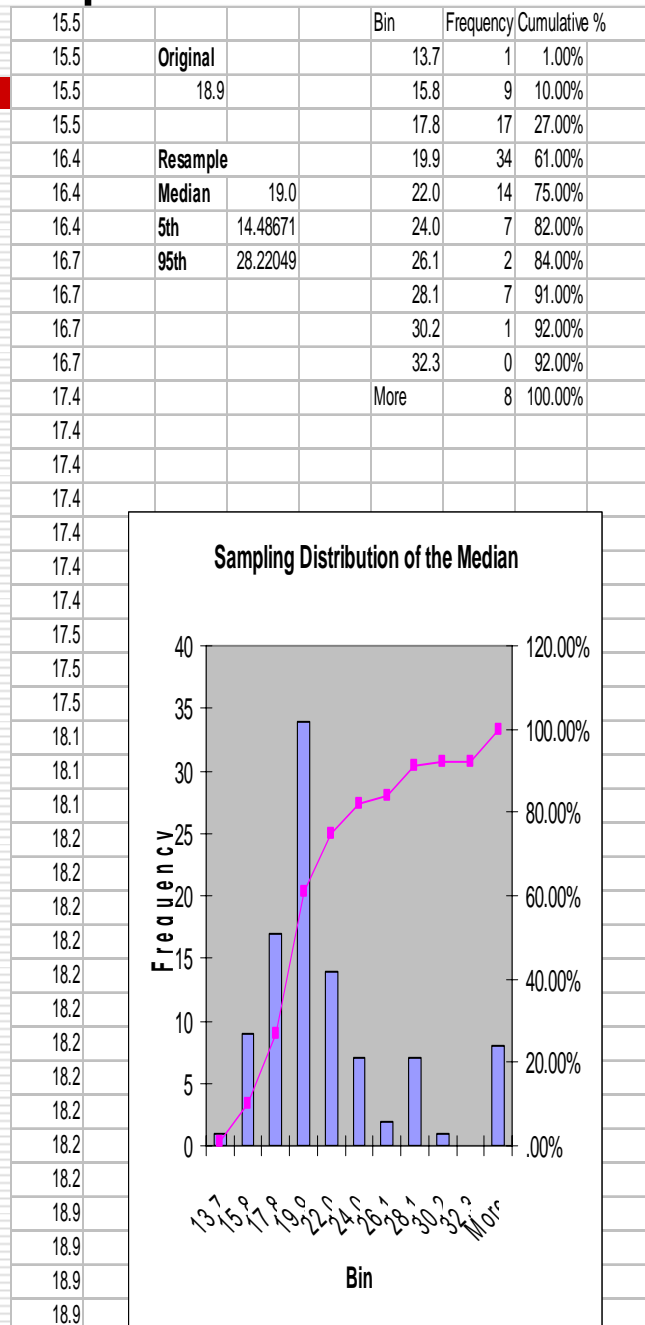




# Distribution of other Statistics from the sample



Baron Von Munchhausen



- There are combinatoric ways to solve for the sampling distribution of the median, but not with normal-theory methods.

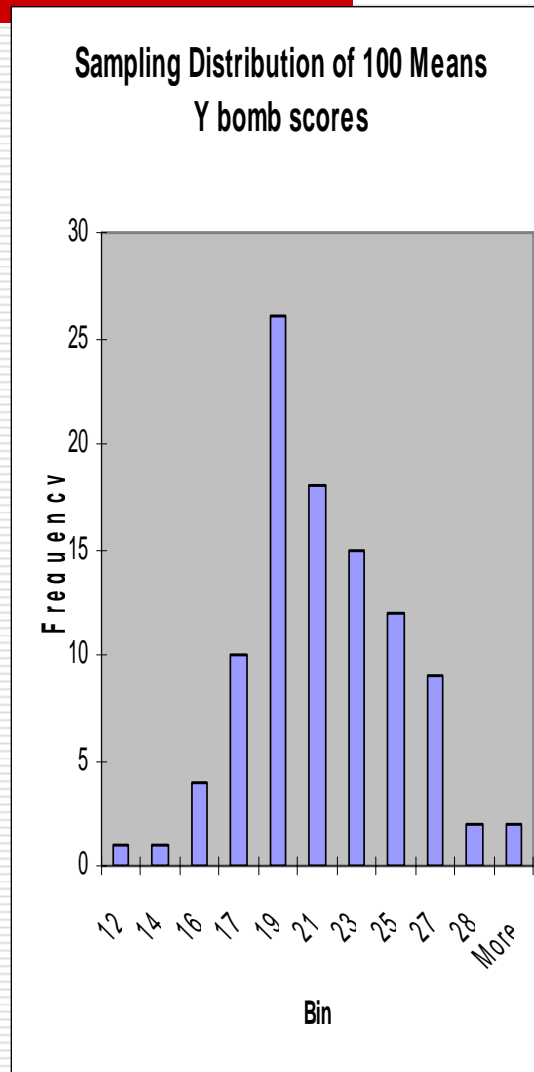




# Sampling distributions enable us to judge *Significance*



Baron Von Munchhausen



- What is the likelihood that the true population average is 8? 19? 35?



Baron Von Munchhausen

# 1 sample of proportions

---



- ☐ Reliability can be viewed as binomial – in any mission, it fails or does not
- ☐ JASSM reliability was one of the motivating problems
- ☐ We worked a sample size (power) problem with a 20 missile launch



Baron Von Munchhausen

## One sample of means, medians, percentiles

---



- ☐ We used the JPF as a motivating example
- ☐ Does the test evidence support the claimed reliability?



Baron Von Munchhausen

## Two sample bootstrap & shuffle



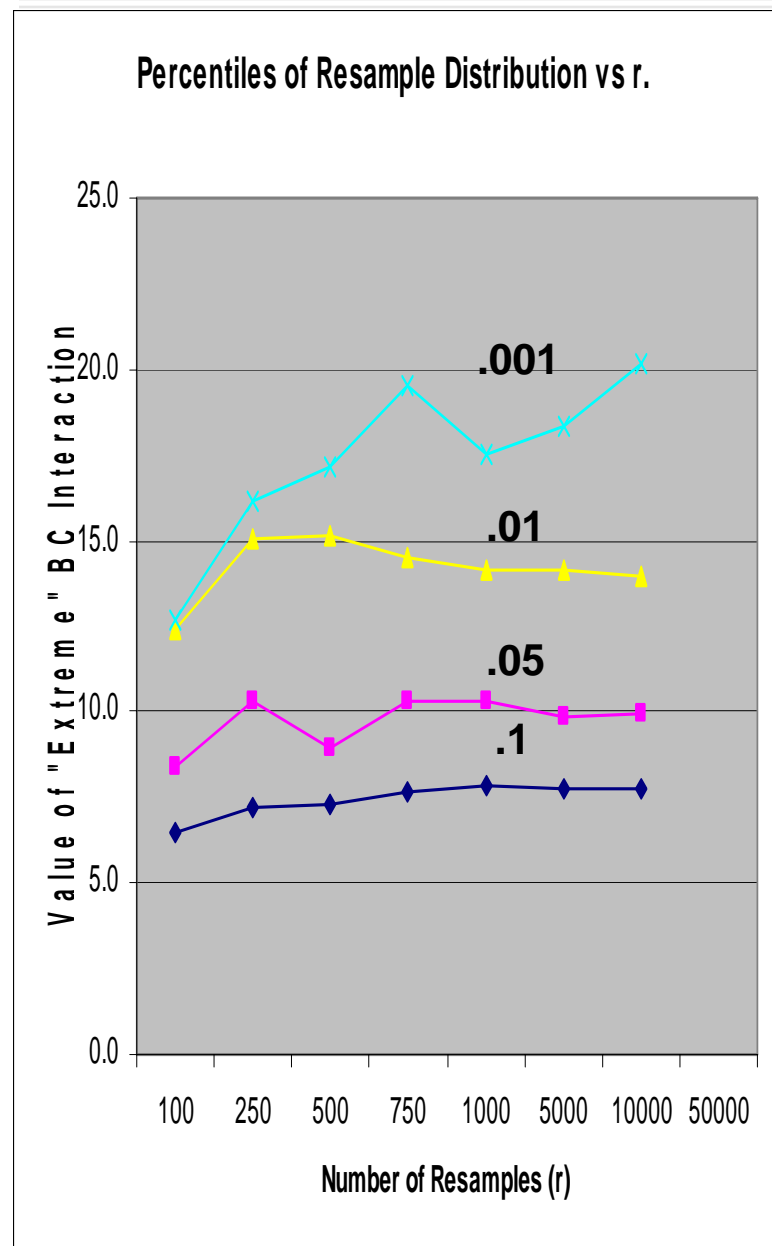
- ☐ We used the "SWEAT Diet" as one of our motivating examples
- ☐ Question 1: do the diets differ in weight lost?
- ☐ Question 2: if so, by how much?



Baron Von Munchhausen

# Resampling Poses Three Sample Size Problems

- How many samples to draw from original data ( $b$ )
  - Best guidance –  $O(n)$  the original sample but NLT 10-15
- How many times to resample original data ( $r$ )
  - Best guidance --  $>500$  times. 1000 converges most problems thus far
- How many original sample data values ( $n$ )
  - Next slide



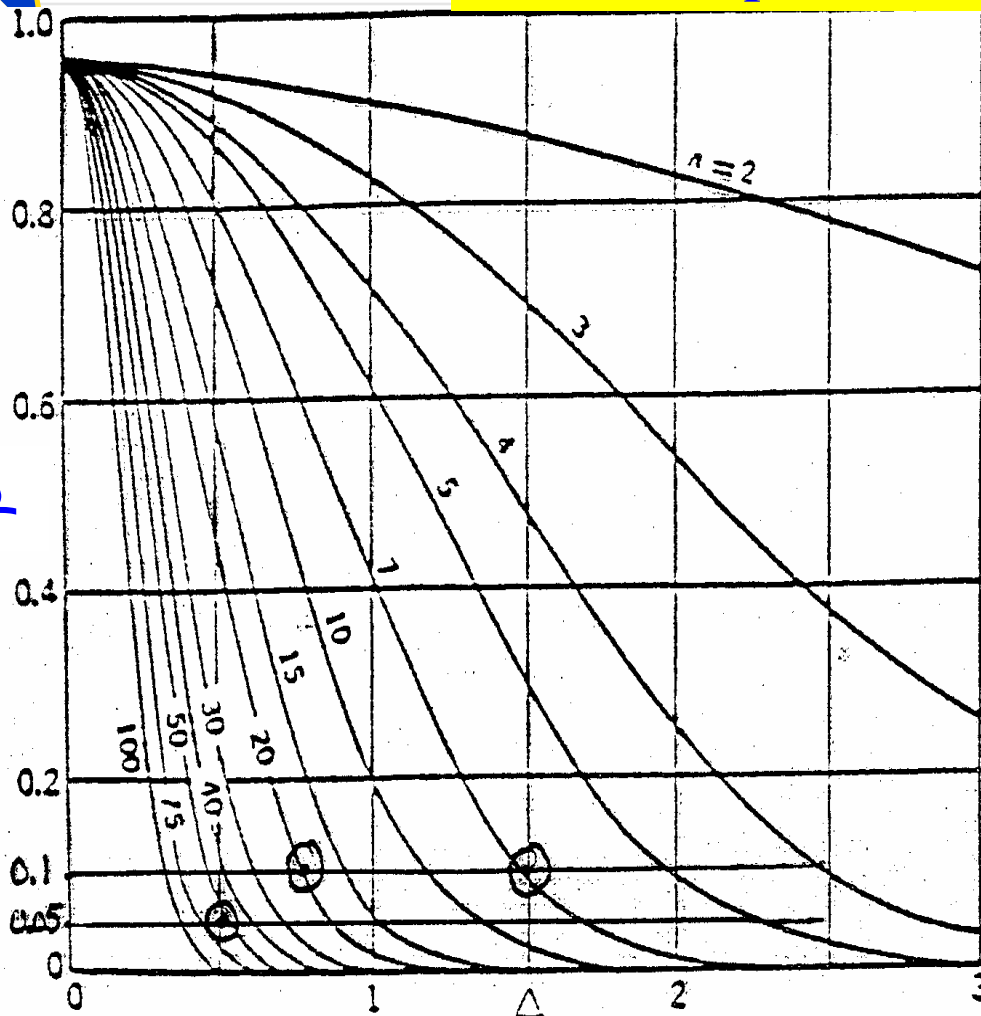


# t Test Sample Size $\alpha$ and $\beta$



Baron Von Munchhausen

$\beta$



OC curves for different values of  $n$  for the two-sided  $t$  test for a level of significance  $\alpha = 0.05$

## 1 Sample

RWR AO  $\leq 12$  sec  $\alpha_1\beta = .05$

$\mu_1 - 12 = 1$  sec

$\sigma = 2$  sec

$\Delta = .05 \Rightarrow n \approx 60$

## 2 Sample (Ind!)

RWR v23 v40 Range diff = 3

$\beta = 0.10$   $\sigma = 2$  miles

$\Delta = .75 \Rightarrow n \approx 20$

$n_1 = n_2 = n^* = \frac{n+1}{2} = \frac{21}{2} = 10.5 \sim 11$

11,  $n_2 = 11$

$$\Delta = \frac{\mu - \mu_2}{2\sigma}$$

$$\Delta = \frac{\mu - \mu_2}{2\sigma}$$

Resamp

trial

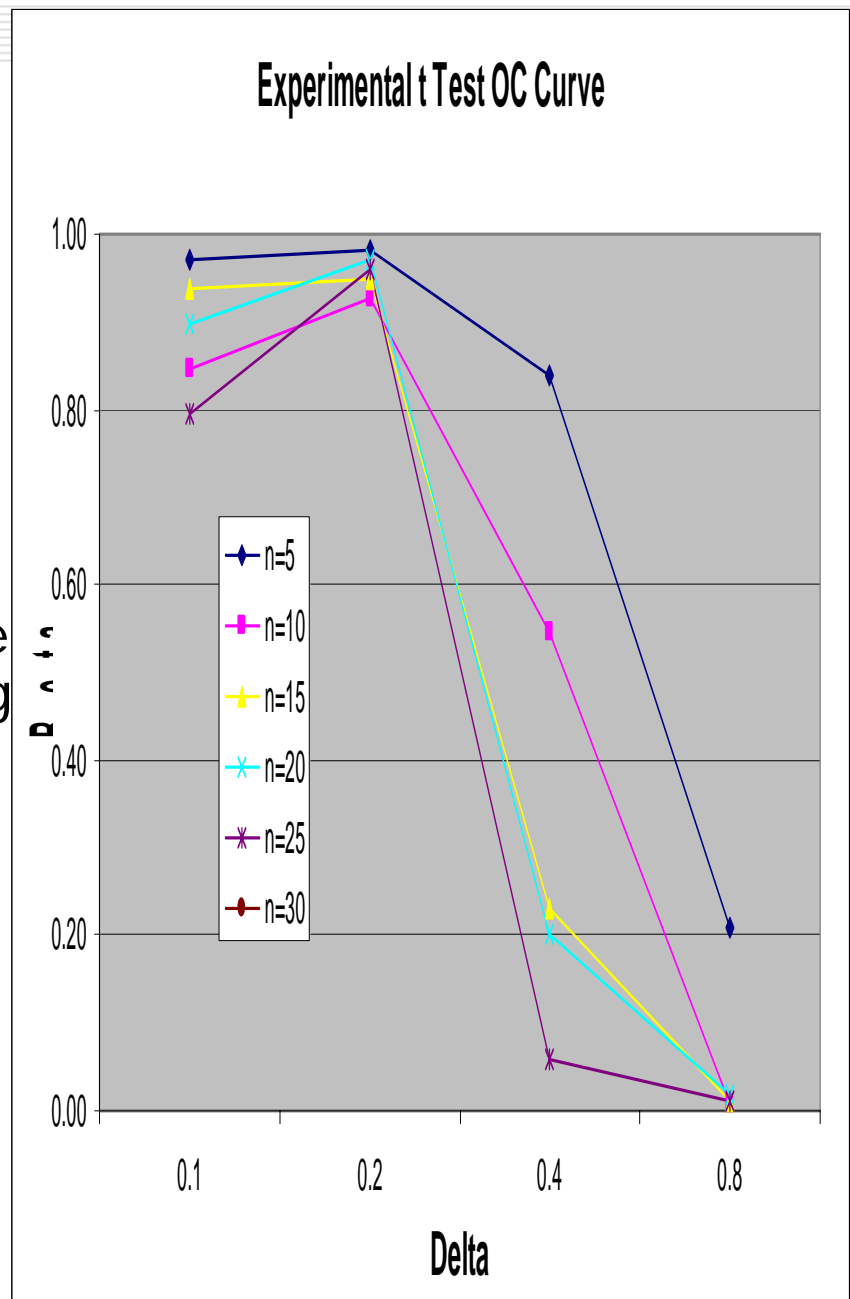


# How many original samples (n)



Baron Von Munchhausen

- As in other areas, resampling clarifies the problem and illuminates the solution
- We more clearly understand the sample size big four:
  - $\alpha, \beta, \delta, \sigma$
- We can replicate the t and F test operating characteristic curves experimentally





# Classical Approaches to Multivariable Testing are Difficult to Master



Baron Von Munchhausen

- ☐ Consider the JSOW delivery:
  - Three terminal delivery parameters
  - Several threat conditions
  - Two profiles
  - Two target types (fixed, mobile)
  - Two sets of launch conditions (in range-in zone)
  - Number of combinations?
- ☐ Our problems are multi-variable
- ☐ So must be our solutions
  - ANOVA
  - Regression
  - Resampling



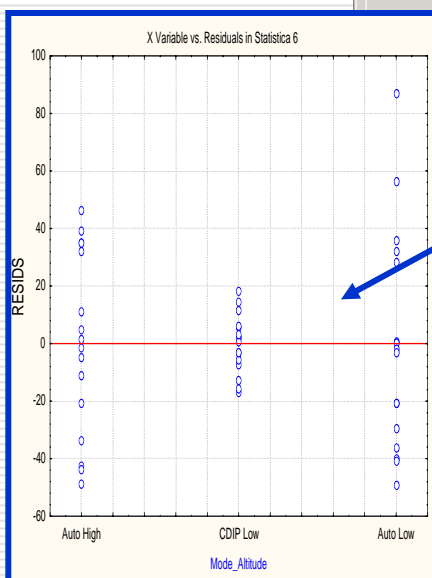
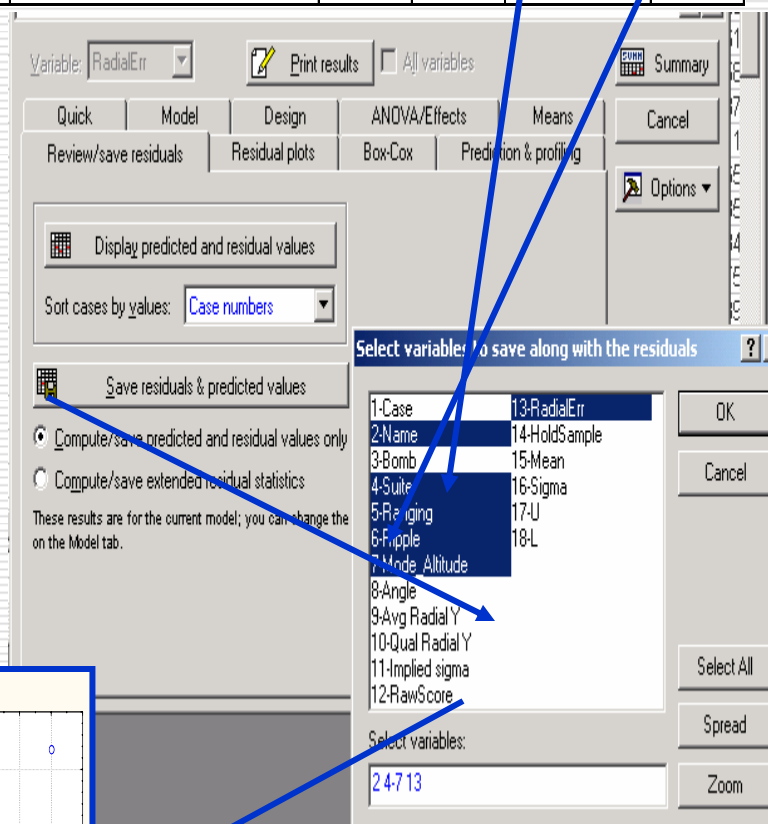


# Compare to simply checking ANOVA assumptions



Baron Von Munchhausen

	Classical Assumptions	Tests			
		NPP	e vs. t	e vs. x, other	e vs. y
1	samples independent of t,x,y ...		✓	✓	
2	equal variances in all cells	✓		✓	✓
3	residuals distributed normally	✓			
4	model is adequate for process		✓	✓	✓



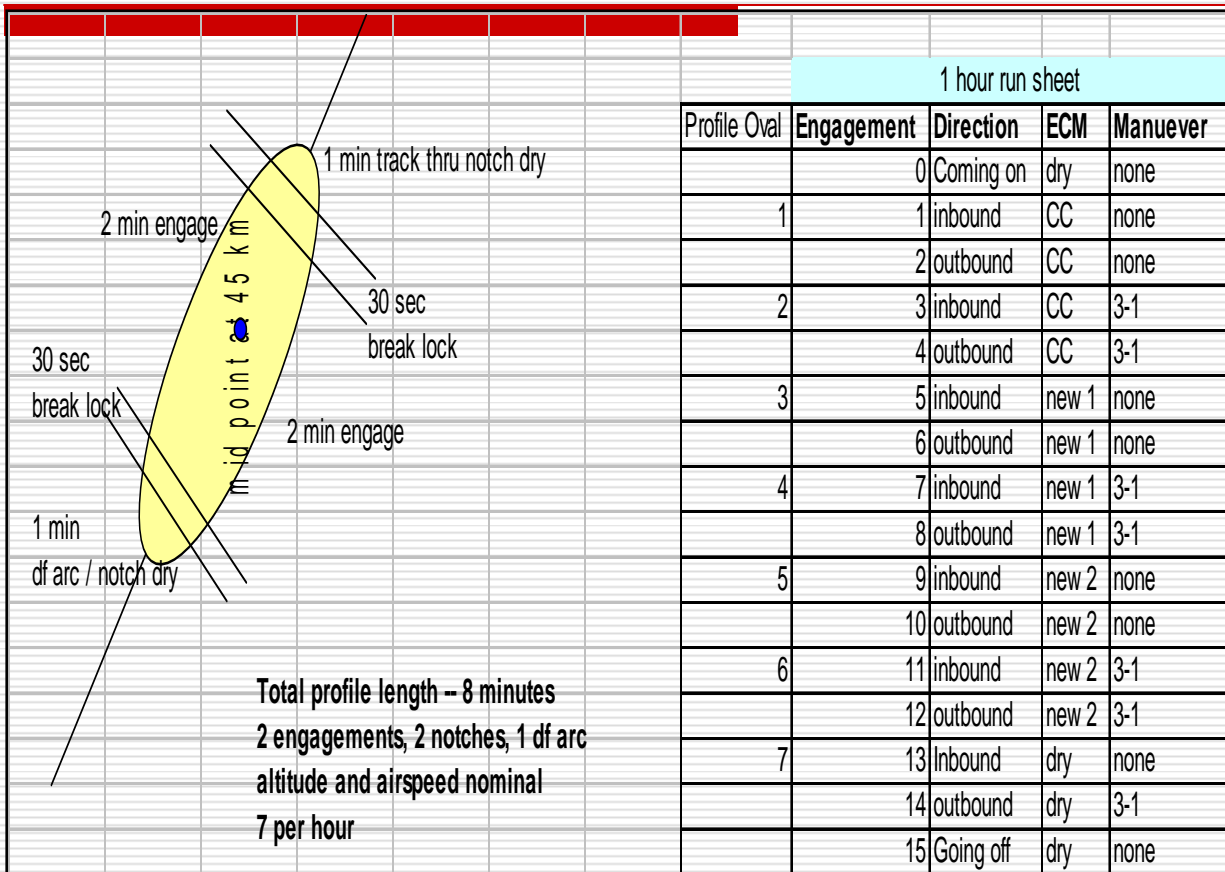
- ❑ On *Review/Save Residuals* tab, choose and save all factors you want to plot against residuals
- ❑ On resulting scroll sheet, do scatter plots with residuals as Y and X or time, or other variable as X



# We're teaching our testers to create multivariable tests



Baron Von Munchhausen



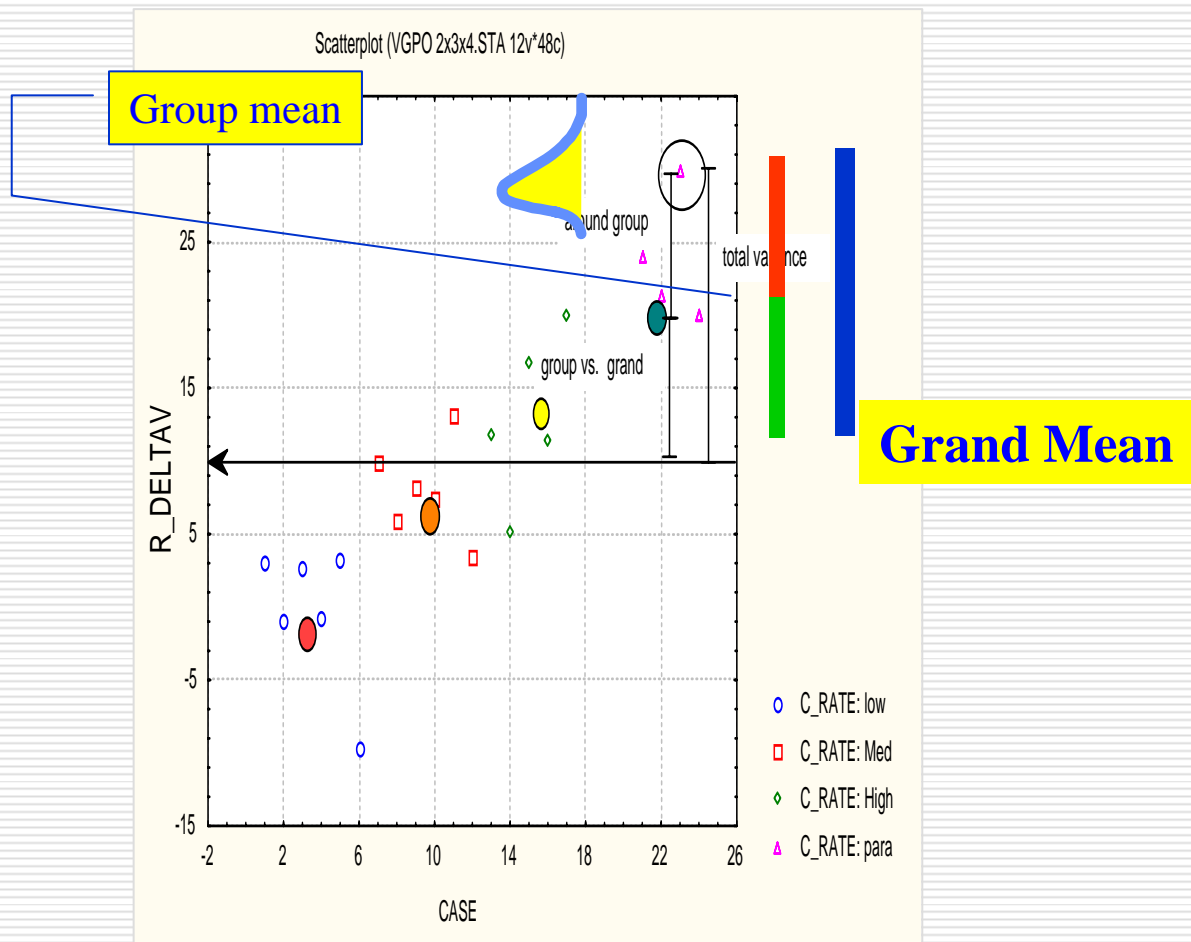
- ☐ Factorial experiment in Mobcap  
Apex threat exploitation examines
  - 4 ECM techniques,
  - two maneuvers, and
  - two directions
- ☐ All in a single one hour mission



# ANOVA partitions variability around the group means



Baron Von Munchhausen



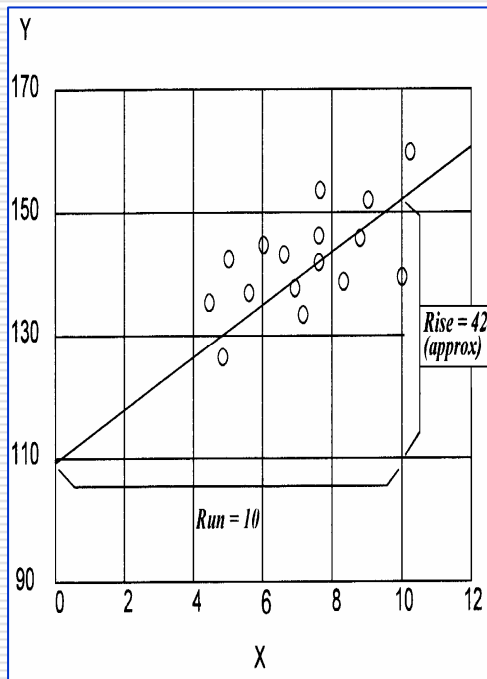
- ANOVA decomposes total variation into explained and unexplained portions



# Regression Partitions Variance around line (not grand mean as ANOVA)



Baron Von Munchhausen



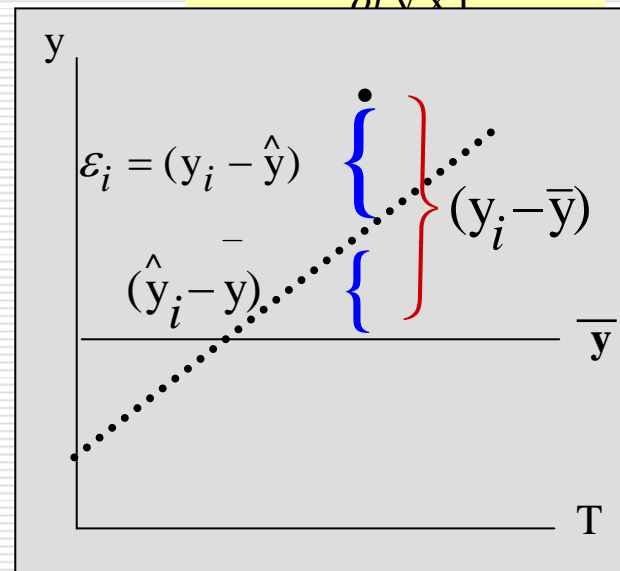
## Assumptions

$$y = \beta_0 + \beta_1 x + \varepsilon$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$\varepsilon_{ij}$  independent

of  $x$  &  $t$



## Regression Coefficients

$$\beta_1 = \frac{\text{Rise}}{\text{Run}} = \text{Slope} = \frac{\Delta y}{\Delta x}$$

$$\beta_0 = \bar{y} - b_1 \bar{x}$$

Recall  $S_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$

## Summing the Squares of the Components of Variance

$$(y_i - \bar{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SST = SSE + SSR$$



Baron Von Munchhausen

# Applying Resampling to DOE

- Sample case had a two variable design with an overall mean of 40 and an A effect of 60 in the presence of 15 units noise
- Resampling Model estimated as follows (90% confidence interval)

	A	B	AB	1	2	3	4	Y Bar
(1)	-1	-1	1	34.7	11.9	16.6	4.9	17.0
b	-1	1	-1	22.5	2.9	13.2	1.6	10.1
a	1	-1	-1	97.8	69.4	79.5	72.3	79.8
ab	1	1	1	101.9	47.2	95.7	84.8	82.4
Effects	67.5	-2.2	4.8					47.3

- This case shows that resampling gives similar answers to classical methods Only A is active in changing Y; B and AB are inactive (effect CI's includes 0)



Baron Von Munchhausen

# Resampling vs. Classical ANOVA

Statistica Results in ExpDesign			Effect CI	
	ANOVA Effect	Std.Err.	-90%	+90%
Mean	47.3	4.0	40.2	54.4
(1)A	67.5	8.0	53.3	81.8
(2)B	-2.2	8.0	-16.4	12.1
1 by 2	4.8	8.0	-9.4	19.1

- We get substantially identical results with resampling and ANOVA
- The algorithm is simple and resistant to “misuse”

## Resampling Results (Within Rows)

	L	U	Estimate	"Real"	CI Width
A	56.2	79.1	67.5	60.0	22.9
B	-13.7	8.3	-2.2	0.0	
AB	-7.4	15.6	4.8	0.0	



# Warning – the Akerson<sup>1</sup> Smear Effect



Baron Von Munchhausen

- ❑ Suppose I have two strong effects  
– A at 60 and AB at 20
- ❑ Estimate these effects via shuffling

Case	A	B	AB	Y				Y Bar
				1	2	3	4	
(1)	-1	-1	1	24.1	19.6	29.7	49.4	30.7
b	-1	1	-1	-24.3	-29.3	-1.6	-8.0	-15.8
a	1	-1	-1	32.4	57.8	34.4	80.0	51.2
ab	1	1	1	85.6	61.0	69.2	70.5	71.6
Effects	53.9	-13.0	33.5					34.4

- When looking up the values in the resampling distribution
  - A of 53.9 is at the 99.4 percentile – a strong effect
  - AB=33.5 is at the 87.5 percentile – an insignificant value
- What happened? The strong A effect is smeared over the rest of the data set, masking the AB effect. Permutation fails here.
- When resampling *within rows* (to keep A where it belongs)
  - AB=33.5 is at the 99.9<sup>th</sup> percentile where it belongs



# Resampling From Non-Normal Data



Baron Von Munchhausen

- ☐ 30 years experience teaches:  
"When resampling and classical methods are applied to classical problems they agree. When applied to no-classical problems, resampling usually gets the right answer."
- ☐ Procedure:
  - Create and examine data that violates normality & equal variance assumptions for ANOVA
- ☐ Findings:
  - For Boot Strap and transformed-ANOVA confidence intervals will not at all agree
  - Boot strap can other useful distribution statistics--including median and percentile





Baron Von Munchhausen

# Some cautionary notes on resampling non-normal data

---

1. There is a bias in bootstrap estimates of variance --, in Confidence intervals -- therefore in significance tests.
2. Bias can be fixed using a second bootstrap procedure.
3. Normal assumptions turn out to be pretty good until the data get way off. Strongly bivariate data are the case we'll show.
4. So far, large skewness doesn't appear to be a major factor.
5. Regardless of this bias, there are numerous important roles for bootstrap.

Warning for beginning users -- dangerous curves ahead.



Baron Von Munchhausen

# Mechanics of analysis via Resampling are simple

---

- ☐ Algorithm:
  - Set up a (possibly constructive) sample
  - Compute a function of the sample observations (Random Variable)
  - Choose a resampling method
    - ☐ Permutation (shuffle)
    - ☐ With replacement (Bootstrap)
    - ☐ With single item deletion (Jackknife)
  - Resample the original sample
  - Review the output for the distribution of the Random Variable
  - Draw conclusions



Baron Von Munchhausen



# Summary and Questions



Baron Von Munchhausen

# Summary of Resampling

- 53<sup>rd</sup> Wing power and confidence across broad battlespace as test method
- Real world poses challenges to classical assumptions
- Resampling provides a method (potentially) that is:
  - Simple to teach and use
  - Resistant to misuse
  - Robust against violations of assumptions
  - Helps with missing data
  - Dealing directly with the problem
  - Illustrative of difficult statistical concepts
- Progress report after 5 years of investigation – very promising – but Akerson Smear a troublesome issue

*“To call in the statistician after the experiment is ... asking him to perform a postmortem examination: he may be able to say what the experiment died of.”*

*Address to Indian Statistical Congress, 1938.*



DOE Founder

Sir Ronald A. Fisher



Baron Von Munchhausen

# Resampling Resources

---

- ❑ Peter Bruce, Resampling Stats, Inc. [pbruce@resample.com](mailto:pbruce@resample.com), [www.resample.com](http://www.resample.com), the single best resource to learn these methods
- ❑ *Resampling: the New Statistics*, Simon and Bruce
- ❑ *Bootstrap Methods*, Michael Chernick
- ❑ *Permutation Tests*, Phillip Good
- ❑ R: A statistics language, freeware for download – numerous sites
- ❑ [Gregory.hutto@eglin.af.mil](mailto:Gregory.hutto@eglin.af.mil)



Baron Von Munchhausen

---

# Questions ? ? ?